

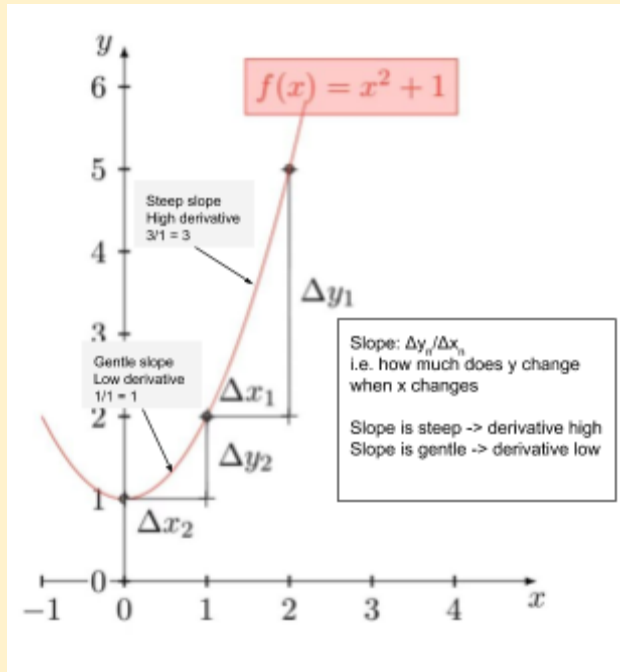
PadhAI: Variants of Gradient Descent

One Fourth Labs

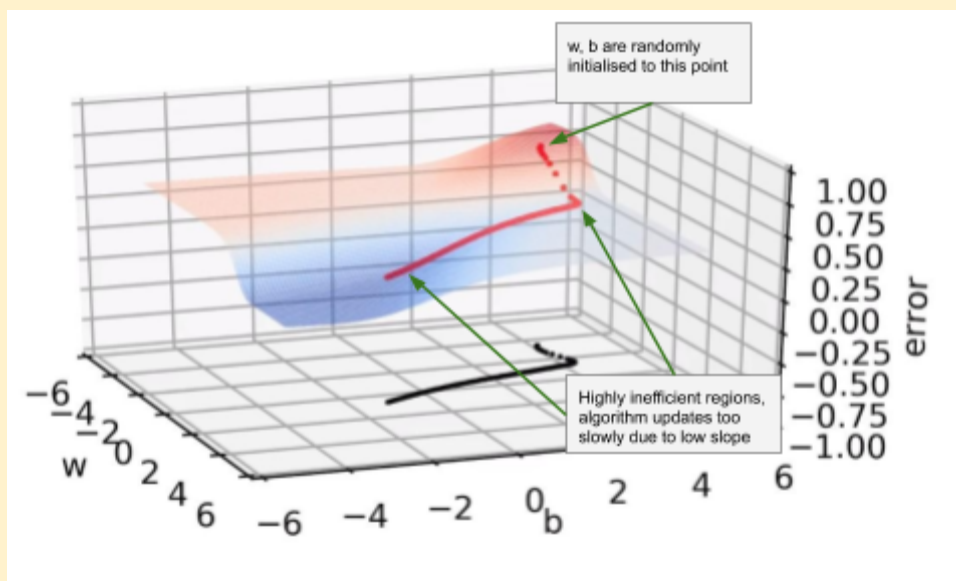
A deeper look into the limitation of Gradient Descent

Why is the behaviour different on different surfaces

1. First, let us look at function to illustrate how the slope behaves



2. So if the derivative is small, the amount by which w or b will be updated by is also small and vice versa if the derivative is large.
3. This could become a problem as in the low-slope/flatter regions, the algorithm does not move fast enough. Here is an example of Gradient descent running inefficiently, moving too slowly.



4. If it so happens that our random initialisation of w and b start at a flat region, then the algorithm would need to run many epochs to get out of the plateau. We need to find a solution for dealing with low slope regions.